A Multimode Anomaly Detection Method Based on OC-ELM for Aircraft Engine System

Chen, Shaowei; Wu, Meng; Wen, Pengfei; Xu, Fangda; Wang, Shengyue; Zhao, Shuai

Published in:
IEEE Access

DOI (link to publication from Publisher):
10.1109/ACCESS.2021.3057795

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE ON IEEE RELIABILITY SOCIETY SECTION

Received January 2, 2021, accepted January 28, 2021, date of publication February 8, 2021, date of current version February 24, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3057795

A Multimode Anomaly Detection Method Based on OC-ELM for Aircraft Engine System

SHAOWEI CHEN[1], (Member, IEEE), MENG WU[1], PENGFEI WEN[1], (Graduate Student Member, IEEE), FANGDA XU[1], SHENGYUE WANG[1], AND SHUAI ZHAO[2], (Member, IEEE)

[1]School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China
[2]Department of Energy Technology, Aalborg University, 9220 Aalborg, Denmark

Corresponding author: Shuai Zhao (szh@et.aau.dk)

ABSTRACT The practical industrial processes possess the characteristics of multimode, unbalanced data distribution, and complex types of abnormalities, which are challenging to the anomaly detection task of complex industrial systems. In this paper, a novel anomaly detection framework based on one-class extreme learning machine (OC-ELM) for the multimode system is presented. To tackle the multiple operation modes, a clustering algorithm is first applied to distinguish the operation modes of the system. The corresponding detection models are built under different operation modes resulting in the multiple models operated in parallel. In addition, the proposed method constructs the reasonable boundary of the complex data distribution, reflecting the equipment running in the healthy or the normal state. The anomaly detection index is obtained according to the deviation degree between the testing sample and the normal model. As a result, a global monitoring index reflecting the degradation state is obtained by combining the anomaly monitoring indices of the equipment under multiple operation modes. The proposed method is verified on a public dataset of aircraft engines, and the advantages are demonstrated by comparing with the implemented detection model without handling the information of operation modes, and the multiple principal component analysis method.

INDEX TERMS Aircraft engine system, anomaly detection, global monitoring index, multiple operation modes, OC-ELM.

NOMENCLATURE

$\boldsymbol{\beta}_i$	output weights of the i^{th} hidden neuron;
$\boldsymbol{H}$	output matrix of the hidden layer;
$\boldsymbol{H}^+$	Moore-Penrose generalized inverse of $\boldsymbol{H}$;
$\boldsymbol{s}$	a testing sample;
$\boldsymbol{w}_i$	input weights of the i^{th} hidden neuron;
$\boldsymbol{x}_j$	j^{th} input vector of the network;
$\boldsymbol{Y}$	output matrix of the network;
$\boldsymbol{y}_j$	j^{th} output vector of the network;
λ	model parameter of the OC-ELM;
μ	fault tolerance of the sample set;
θ	predefined threshold of the anomaly index;
ε	accepted minimum value;
$d(\cdot)$	distance function;
$f_L(\cdot)$	output function of the network;
$g(\cdot)$	activation function of the network;
L	number of the neurons on hidden layer;
N	number of the distinct samples;

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaojun Li.

I. INTRODUCTION

The main task of the industrial process monitoring is to monitor the changes of the system states constantly and timely so as to capture the anomaly state before the system failure, which can help to avoid serious damages to the system [1], [2]. At present, the research work on anomaly detection is conducted commonly under the assumption of the single operation mode, i.e., the system is within the same operation mode during its whole operation process. In this case, a model is established to analyze the normal and abnormal behaviors of the system [3]–[5]. However, in practical industrial processes, the operation mode of the equipment usually varies along with the changes of productive conditions, environment

and so on, in which the industrial equipment undergoes different operation modes throughout its operation process. Since the detection model based on single operation mode cannot detect all types of anomalies under multiple operation modes, traditional anomaly detection methods are no longer applicable in such cases. They may ignore the local behavior of the system, missing valuable information. Through the process monitoring under multiple operation modes, more comprehensive anomaly detection can be realized to ensure the safety and the reliability of the system.

For a system undergoing multiple operation modes, the information of normal samples in each operation mode is different. It is inaccurate to establish a global detection model in this case, in which the output deviation caused by the change of operation modes may be falsely identified as the system anomaly behavior. By contrast, multiple models established simultaneously can accurately characterize the behavior of the system under each operation mode [6]. Establishing multiple models not only reduce the complexity of the data set, but also facilitate timely capturing the characteristic changes of the data. By quantitatively evaluating the healthy state of the system, the optimal timing for equipment maintenance is determined, so that a reasonable maintenance strategy can be developed as early as possible to improve the availability of equipment. For the practical industrial process with multiple operation modes, the label characterizing the current operation mode is commonly unknown. In fact, collecting the labels by traversing all of the operation modes of the equipment is usually time-consuming and costly. In addition, the industrial equipment is generally in a normal operation stage, and the probability of failure is small, resulting in less samples of anomaly. It is unrealistic to get sufficient anomaly data of the device in all abnormal situations. The above factors make it difficult to identify the health state of the system. Consequently, it is particularly important and challenging to achieve multimode anomaly detection for the industrial process.

With the development of the sensor network technology, many multivariate monitoring methods have been successfully applied to the process monitoring [7], [8]. Among them, principal component analysis (PCA) is one of the most applied multivariable monitoring method in the industrial process monitoring [9], [10]. PCA is an unsupervised algorithm to realize anomaly detection without knowing the labels of the data, which makes it suitable for the practical industrial process [11]. However, PCA also has limitations, e.g., it is assumed that the data follow a Gaussian distribution. Measurements from the actual industrial process are often non-Gaussian, non-linear, and more complicated. At present, there is some literature [12] in which the process monitoring problems are divided as the one-class classification problems, which identify the system anomaly by learning the boundaries between the normal and abnormal data. For example, support vector data description (SVDD), one-class extreme learning machine (OC-ELM), local outlier factor (LOF), and one-class support vector machine (OC-SVM) can well characterize changes of non-Gaussian and non-linear data. In [13], OC-ELM without adjusting the hidden layer and the output weights was proposed. In the presence of the high dimensionality and a large amount of the aeronautical data, Janakiraman and Nielsen [14] used OC-ELM to achieve anomaly detection with the high-speed training and the good performance of generalization. Imamverdiyev and Sukhostat [15] proposed a network intrusion detection method based on ELM. In [16], OC-ELM was applied to detect anomalies of the gas turbine combustor. Mygdalis *et al*. [17] proposed the laplacian OC-ELM for the recognition of human action. In [18], OC-ELM was utilized to achieve the intelligent video analysis. Although OC-ELM has been well applied in anomaly detection in recent years [19], there are few papers about the utilization of OC-ELM with multiple operation modes.

Multimode anomaly detection is one of the relevant topics in the field of anomaly detection. In [20], fault detection for the suspension system of the maglev train under multiple modes was studied. Bakdi *et al.* [21] combined static PCA and dynamic PCA to monitor modern wind turbines. Yang *et al.* [6] developed a robust dictionary learning method for the multimode aluminum electrolysis process. Wang *et al.* [22] realized the multimode process monitoring based on LOF for the benchmark Tennessee Eastman process. Wang *et al.* [23] proposed a multisubspace factor analysis (FA) method, and the SVDD model was constructed to identify faults. In [9], a multiscale neighborhood normalization-based multiple dynamic PCA method was proposed for a ladle furnace steelmaking process case. At present, most of studies on multimode anomaly detection focus on chemical processes, which may lack an universal applicability. The research and development on the multimode anomaly detection technology for other industrial equipment such as aircraft engines is few and not mature. In addition, it is not necessary for OC-ELM to involve anomaly samples to build models, which is able to achieve good effects of anomaly detection in different fields.

In this paper, the multimode anomaly detection of aircraft engines based on OC-ELM (Multi-OC-ELM) is developed and implemented. Regarding the anomaly detection under multiple operation modes, a novel framework is developed for the aircraft engine system. The case of multimode studied in this paper is more complicated, i.e., the corresponding operating condition at each time step is randomly given, indicating a more frequent change of the operation modes compared with common industrial processes. First, multiple operation modes of the aircraft engine are separated utilizing a clustering algorithm. Then, under each operation mode, OC-ELM is adopted to construct the detection model. Finally, the anomaly monitoring indices can be obtained to indicate the anomaly degree of the equipment. To illustrate the effectiveness and the superiority, the proposed method is compared with the sole application of OC-ELM (Single-OC-ELM) model and the multimode PCA model (Multi-PCA). As a result, the proposed method is able to effectively deal with the

multimode case of aircraft engines. The main contributions of this paper are summarized as follows:

(1) This paper puts forward an anomaly detection framework based on OC-ELM to mitigate the multiple working condition issue. A clustering algorithm is applied to split data under multiple working conditions of aircraft engines. In this way, it mitigates the challenges of detecting anomalies of systems with multiple working conditions and eliminate the possible detection deviation and bias.

(2) The data collected in the actual industrial systems are usually unbalanced since the abnormal samples are always difficult to be obtained in real applications. In the proposed method, there is no need to collect abnormal samples to build the model, which can achieve generalizable anomaly detection and lower the data requirements in different fields.

(3) The proposed method is generic and can realize a more comprehensive and accurate anomaly detection of other complex industrial systems. As an illustrative example, the effectiveness of the proposed method is verified on the C-MAPSS dataset.

The rest of this paper is organized as follows. Section II provides the framework of multimode anomaly detection and describes the algorithm principle. In Section III, a case study of the aircraft engine is carried out, and the comparison with other methods demonstrates the superiority of the proposed method. Finally, the conclusions are summarized in Section IV.

II. METHOD DEVELOPMENT

A. ALGORITHM PRINCIPLES

1) EXTREME LEARNING MACHINE

In [24], Huang *et al.* proposed ELM based on the single-hidden layer feedforward neural networks (SLFNs). The ELM network has only one parameter that requires to be tuned. The input weights and the hidden layer biases are randomly initialized. Then, the output weights can be obtained by solving the linear system of equations.

For the SLFNs shown in Fig. 1, assume that the number of the hidden layer neurons is L, and the output function of the network $f_L(\boldsymbol{x})$ can be represented as

$$f_L(\boldsymbol{x}) = \sum_{i=1}^{L} \boldsymbol{\beta}_i g(\boldsymbol{\omega}_i, b_i, \boldsymbol{x}), \tag{1}$$

where $g(\boldsymbol{x})$ is the activation function determining the relationship between the input vector $\boldsymbol{x}$ and the i^{th} hidden layer neuron. Suppose that there are N arbitrary distinct samples $(\boldsymbol{x}_j, \boldsymbol{y}_j)$, $j = 1, 2, \cdots, N$, and the input vector is $\boldsymbol{x}_j = [x_{j1}, x_{j2}, \cdots, x_{jn}]^T \in R^n$. The output vector of the network is $\boldsymbol{y}_j = [y_{j1}, y_{j2}, \cdots, y_{jm}]^T \in R^m$. Thus, the output of the SLFNs can be represented as

$$\sum_{i=1}^{L} \boldsymbol{\beta}_i g(\boldsymbol{\omega}_i \cdot \boldsymbol{x}_j + b_i) = \boldsymbol{y}_j, \quad j = 1, 2, \cdots, N, \tag{2}$$

where $\boldsymbol{\omega}_i = [\omega_{i1}, \omega_{i2}, \cdots, \omega_{in}]^T$ are the input weights and $\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \cdots, \beta_{im}]^T$ are the output weights of the i^{th} hidden layer node. b_i is the bias of the i^{th} hidden layer node and $\boldsymbol{\omega}_i \cdot \boldsymbol{x}_j$ represents the inner product of $\boldsymbol{\omega}_i$ and $\boldsymbol{x}_j$.

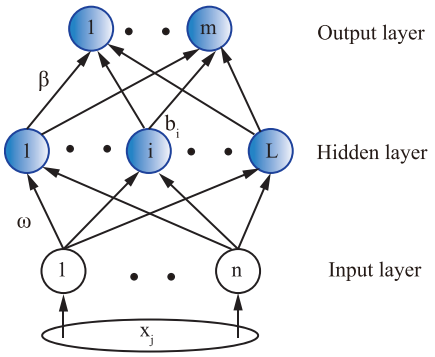


FIGURE 1. Structure of the ELM network.

The corresponding matrix form can be written as

$$\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{Y}, \tag{3}$$

where

$$\boldsymbol{H}(\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_L, b_1, \cdots, b_L, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_N) = \begin{bmatrix} g(\boldsymbol{\omega}_1 \cdot \boldsymbol{x}_1 + b_1) & \cdots & g(\boldsymbol{\omega}_L \cdot \boldsymbol{x}_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(\boldsymbol{\omega}_1 \cdot \boldsymbol{x}_N + b_1) & \cdots & g(\boldsymbol{\omega}_L \cdot \boldsymbol{x}_N + b_L) \end{bmatrix}_{N \times L},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_L^T \end{bmatrix}_{L \times m}, \quad \boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_1^T \\ \vdots \\ \boldsymbol{y}_N^T \end{bmatrix}_{N \times m}.$$

$\boldsymbol{H}$ is the output matrix of the hidden layer and $\boldsymbol{Y}$ is the network output. In most cases, the number of the hidden layer neurons is less than the number of samples, i.e., $L \ll N$, and it can be obtained as

$$\| \boldsymbol{H}_{N \times L} \boldsymbol{\beta}_{L \times m} - \boldsymbol{Y}_{N \times m} \| < \varepsilon, \tag{4}$$

suggesting that the error of the network can reach an accepted minimum value ε by learning the given samples. The least-squares solution $\hat{\boldsymbol{\beta}}$ is solved as

$$\min_{\hat{\beta}} \| \boldsymbol{H}\hat{\boldsymbol{\beta}} - \boldsymbol{Y} \|, \tag{5}$$

which is equivalent to minimizing the loss function

$$E = \sum_{j=1}^{N} \left(\sum_{i=1}^{L} \boldsymbol{\beta}_i g\left(\boldsymbol{\omega}_i \cdot \boldsymbol{x}_j + b_i\right) - \boldsymbol{y}_j \right)^2. \tag{6}$$

Hence, when the activation function is given and the values of ω and b are randomly assigned, the output matrix $\boldsymbol{H}$ becomes a constant matrix, in which the whole network is simplified as a linear system. $\boldsymbol{\beta}$ is determined by

$$\hat{\boldsymbol{\beta}} = \boldsymbol{H}^{+}\boldsymbol{Y}, \tag{7}$$

where $\boldsymbol{H}^{+}$ is the Moore-Penrose generalized inverse of the matrix $\boldsymbol{H}$. It can be proved that $\hat{\boldsymbol{\beta}}$ is the unique minimum norm solution.

2) ONE-CLASS EXTREME LEARNING MACHINE

ELM can tackle multiclass classification problems, realizing fault diagnosis and pattern recognition. However, the equipment in practical industry is usually in healthy states with a large number of healthy samples. The small probability of the device abnormality leads to the imbalance of two kinds of samples, which is not suitable to use the original ELM for identification. When OC-ELM is used to perform classification, it implements modeling only involving healthy samples without using anomaly samples, which is applicable to the industrial process.

The output function of the ELM network can be written as

$$f(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x})^T \boldsymbol{\beta}, \tag{8}$$

where $\boldsymbol{h}(\boldsymbol{x})$ realizes nonlinear feature mapping of $R^n \rightarrow R^L$. OC-ELM can separate the normal and the abnormal data in the case where only the normal data are available, and it regards the normal samples as the target class to achieve modeling. As a result, β becomes an approximate linear mapping, i.e., a hyper plane approximation. There is a basic assumption that similar objects are close in the feature space producing similar outputs. Therefore, the expected output is supposed to be the same for the target class in the one-class classifier, which is

$$y_j = p, \quad \forall \boldsymbol{x}_j \in \boldsymbol{X}, \quad j = 1, 2, \cdots, N, \tag{9}$$

where $\boldsymbol{X}$ is the overall input sample set, and p is a real number. The expected outputs of all training samples are supposed to be set to the same value p for learning the boundary of the detection model. The expected target output vector is

$$\boldsymbol{Y} = [y_1, y_2, \cdots, y_N]^T = [p, \cdots, p]^T, \tag{10}$$

and the training process is the same as ELM. When a testing sample $\boldsymbol{s}$ arrives, the one-class classifier defines a distance function d, which is

$$d(\boldsymbol{s} \mid \boldsymbol{X}, \lambda) = \mid \boldsymbol{h}(\boldsymbol{s})^T \boldsymbol{\beta} - p \mid, \tag{11}$$

representing the distance from an arbitrary sample point $\boldsymbol{s}$ to the hyper plane. In (11), λ is the model parameter representing the complexity of the model [13], e.g., the number of hidden nodes of ELM, etc. The obtained distance d indicates the anomaly degree of the testing sample. A large value d means that the testing sample is far away from the target class, and thus the more likely it is abnormal. It is supposed to be identified based on a predefined threshold θ. Generally, the threshold θ is obtained from the training process. The distance from training samples to the target class is

$$d(\boldsymbol{x}_j \mid \boldsymbol{X}, \lambda) = \mid \boldsymbol{h}(\boldsymbol{x}_j)^T \boldsymbol{\beta} - p \mid \tag{12}$$

where the obtained distance represents the training error. Then, the distance values of all training samples are denoted as $\boldsymbol{d} = \left[d_{(1)}, \cdots, d_{(N)}\right]$, $d_{(1)} \geq d_{(2)} \geq \cdots \geq d_{(N)}$. Hence, the threshold θ can be written as

$$\theta = d_{\text{floor}(\mu \cdot N)}, \tag{13}$$

where μ represents the fault tolerance of the sample set, mitigating over-fitting and enhancing the robustness of the model. N is the number of training samples. floor(a) returns the largest integer not greater than a. The decision function of the testing sample $\boldsymbol{s}$ is

$$\begin{aligned} D_{\text{OC-ELM}}(\boldsymbol{s}) &= \text{sign}(\theta - d(\boldsymbol{s} \mid \boldsymbol{X}, \lambda)) \\ &= \begin{cases} 1 & \text{Normal class,} \\ -1 & \text{Abnormal class,} \end{cases} \end{aligned} \tag{14}$$

which finally realizes the anomaly detection for testing samples.

Algorithm 1 One-Class Extreme Learning Machine

Input:
Input sample set $\boldsymbol{X}$,
Expected target output vector $\boldsymbol{Y} = [y_1, y_2, \cdots, y_N]^T = [p, p, \cdots, p]^T$,
Testing sample $\boldsymbol{s}$.
Output:
$D_{\text{OC-ELM}}(\boldsymbol{s}) = \text{sign}(\theta - d(\boldsymbol{s} \mid \boldsymbol{X}, \lambda))$.
Process:
1: Assign a random value to $\boldsymbol{\omega}_i$, b_i;
2: Compute the output matrix of the hidden layer $\boldsymbol{H}(\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_L, b_1, \cdots, b_L, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_N)$;
3: Compute the output weights $\hat{\boldsymbol{\beta}} = \boldsymbol{H}^+ \boldsymbol{Y}$;
4: Compute the distance from the training sample to the target output $d(\boldsymbol{x}_j \mid \boldsymbol{X}, \lambda) = \mid \boldsymbol{h}(\boldsymbol{x}_j)^T \boldsymbol{\beta} - p \mid$;
5: Obtain the threshold according to the training error $\theta = d_{\text{floor}(\mu \cdot N)}$;
6: Utilize the decision function $D_{\text{OC-ELM}}(\boldsymbol{s})$ treat the testing sample $\boldsymbol{s}$.

B. FRAMEWORK OF ANOMALY DETECTION UNDER MULTIPLE OPERATION MODES

The whole monitoring process is divided into two phases as shown in Fig. 2, the training phase and the testing phase. In the training phase, operation modes determination and multimodel establishment are conducted. As for the testing sample, its operation mode is determined first, and then it is sent to the corresponding detection model to obtain the corresponding anomaly monitoring index. Finally, the anomaly indices under multiple operation modes are recombined, so as to obtain the global monitoring index of the equipment, reflecting the degradation degree of the equipment under different operation modes with cycles increasing. The abnormal state of the system can be detected based on the anomaly monitoring index by setting the corresponding threshold.

1) TRAINING PHASE

a: OPERATION MODES DETERMINATION

This paper adopts the parallel strategy of multimodel for multiple operation modes. The operation modes are distinguished first. k-means is applied to separate the operation

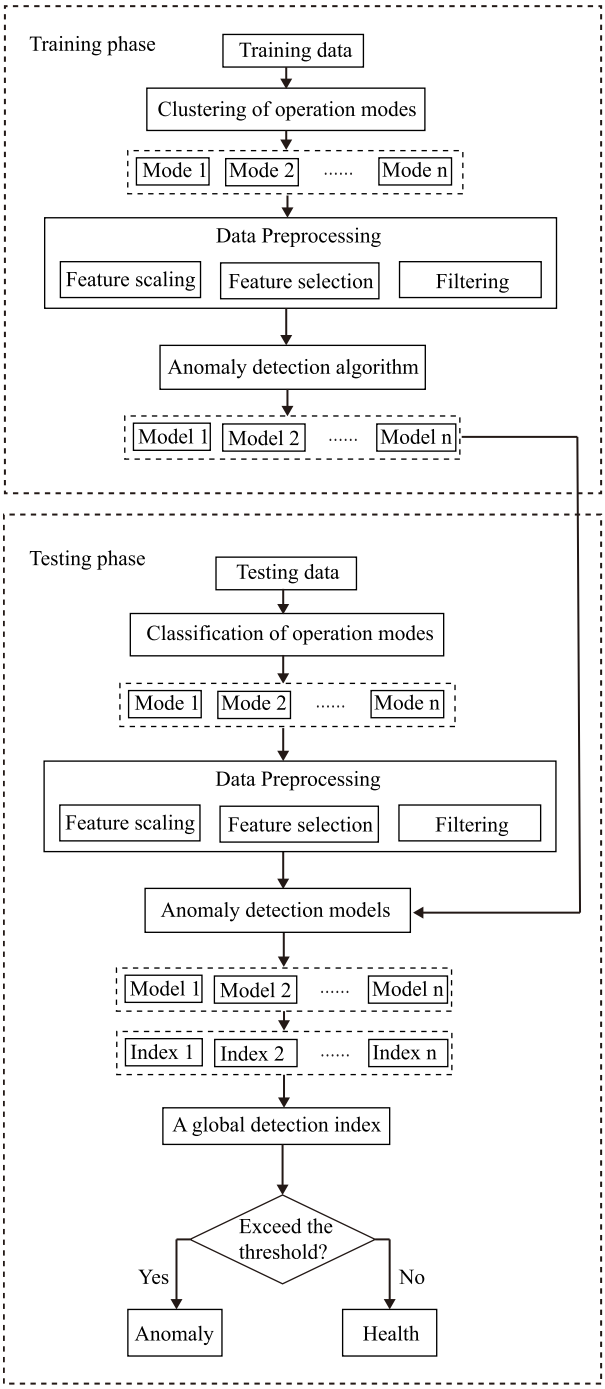


FIGURE 2. Framework of the proposed multimode anomaly detection method.

modes automatically without knowing the labels of the corresponding operation modes. In this paper, the training data set is clustered by k-means, in which each cluster represents an operation mode.

b: DATA PREPROCESSING

Generally, a field dataset is collected by a large number of sensors to jointly describe the operating state of the equipment. However, excessive sensors may result in information redundancy. Moreover, different sensor data have different ranges, and most of the original data are contaminated by noise during the data acquisition process. Therefore, after distinguishing the operation modes, data are preprocessed under each operation mode respectively.

FEATURE SCALING

Feature scaling is implemented to eliminate the influence caused by the different magnitude of different sensors. Data standardization makes all sensor data be within the same magnitude for comparison. In this paper, z-score standardization is employed to obtain standardized data by calculating the mean and variance of the original data [25]. The standardized data follow a normal distribution with the mean 0 and the standard deviation 1. The standardized data x^* can be obtained as

$$x^* = (x - \mu)/\sigma, \tag{15}$$

where μ and σ are the mean and standard deviation of the original data respectively.

FEATURE SELECTION

For the high-dimensional dataset, the data of some sensors may not change with increasing cycles and will provide less valuable information. In addition, a few sensors have different development trends on different devices, which may affect the accuracy of constructing detection models. Sensors with the above characteristics would be removed, and this paper prefers to select those sensors from which the collected data show apparent trends as well as the similar trends for all devices. As a result, the sensors selected under each operation mode are not exactly the same. The sensor selection reduces the dimension of the original data set, thus reducing the complexity of model building and the time cost of model training.

FILTERING

Generally, the actual industry dataset may contain noisy data or outliers. The moving average filtering is adopted as the filtering technique. The mean of the original data within a fixed-length moving window is taken as a new data point, thereby realizing the data smoothing and noise reduction.

c: CONSTRUCTION OF MULTI-OC-ELM MODELS

Under the condition where there are a large number of healthy samples and rare anomaly samples, this paper uses OC-ELM to achieve anomaly detection. OC-ELM constructs the detection model only involving the healthy samples [13]. The anomaly index of the testing sample is obtained by measuring the deviation degree between the testing sample and the normal samples. In this paper, the corresponding detection models are trained respectively using the historical health data under each operation mode, achieving a parallel construction of multiple models.

2) TESTING PHASE

a: IDENTIFICATION OF OPERATION MODES

In the training phase, the labels indicating the corresponding operation modes are determined by using k-means. In the testing phase, ELM is adopted for the identification of operation modes. ELM is trained by using the operation mode data separated by k-means in the training phase, and then the operation mode of the testing sample is classified by the trained ELM model.

b: DATA PREPROCESSING

The data preprocessing here can refer to the previous training part.

c: ANOMALY DETECTION

After the data preprocessing, the anomaly index of the testing sample is obtained based on the trained OC-ELM model under the corresponding operation mode. Since multiple detection models are established, the anomaly indices of the same equipment under different operation modes will be obtained.

d: RECOMBINATION OF MULTIPLE INDICES

With multiple anomaly indices from multiple operation modes, it is expected to get a global monitoring index through recombining the multiple indices to represent the changes of the degradation state during the whole operating process of the system. Usually, as the cycles increasing, the performance of the system will irreversibly degenerate due to the system degradation. Once the system begins to deviate from its healthy state, the degree of abnormality in the system will gradually deteriorate until it fails. Therefore, the changes of the whole degradation trend in the system can be observed by arranging the anomaly indices from multiple operation modes according to the original chronological order of the operation modes. More generally, at each monitoring time, when values of anomaly indices are output by the Multi-OC-ELM model, they were then concatenated in terms of the monitoring time to produce the global anomaly index.

III. METHOD VERIFICATION AND PERFORMANCE ILLUSTRATION

In this paper, a new multimode anomaly detection method based on OC-ELM is proposed, in which there are multiple Multi-OC-ELM models built in parallel. In the training phase, the multimode data set is divided into several subsets according to the identified operation modes. Then the corresponding detection models are established respectively under different operation modes by OC-ELM. When the testing sample arrives, its operation mode is identified first and then it is sent to the corresponding detection model for similarity comparison to obtain the anomaly indicator. Finally, the anomaly indicators from multiple operation modes under the same engine are recombined, which can reflect the change of the anomaly degree during the entire degradation process. In this paper, the proposed methods were verified based on MATLAB 2016B [26].

A. DATASET DESCRIPTION

The data set used in this paper consists of 260 multivariate time series generated by Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) [27], [28]. The C-MAPSS data set is acquired from a commercial simulation software simulating the degradation process of aircraft engines. It contains four subsets, and the subset involving a single failure mode and 6 operating conditions (FD002) is adopted here. This subset consists of a training set (train_FD002) and a testing set (test_FD002), where the former contains 260 run-to failure units and a total of 53,759 condition monitoring measurements. It should be mentioned that only train_FD002 was used in this case because our goal is to detect the anomaly, while there are only partial measurements of each unit were collected when they were in an approximately healthy state in test_FD002. Each multivariate time series represents the progressive degradation of an engine. Each engine is in a random health state at the beginning. As the anomaly occurs and deepens, the engine gradually degenerates until it fails finally.

In the engine data set, the health state of each engine is monitored by 21 sensors. In addition, the operation mode data are composed of the altitude (0-42K feet), mach number (0-0.84), and throttle resolver angle (20-100), indicating the cruise conditions of the aircraft. Moreover, the operation modes of each engine are generated by selecting different cruise conditions at each time step randomly rather than periodic change. The engine data set includes totally six operation modes and the failure of all engines is caused by the degradation of high-pressure compressor (HPC).

B. DATA PROCESSING

The proposed multimode anomaly detection framework for the aircraft engine system mainly contains the training phase and the testing phase. Therefore, the data set is divided into a training set and a testing set. 100 engines are randomly selected to form the training set, and the testing set is composed of the remaining 160 engines.

In the training phase, distinguishing different operation modes in the training set is conducted first. However, it is difficult for users to customize different operation modes in the identification of operation modes of aircraft engines. Despite the fact that there are 6 operating conditions that had been priorly known, we followed some existing researches to cluster the data in terms of their corresponding operating conditions, and 6 clusters were then acquired [29]. The three-dimensional visualization of the operation modes of the aircraft engine is shown in Fig. 3. It can be observed clearly that the data set can be separated into six areas completely, possessing a high degree of aggregation of points in each area. Considering the above factors, k-means is used in this paper to locate six clusters automatically. k is set to 6 in k-means to obtain different sample groups under six operation modes.

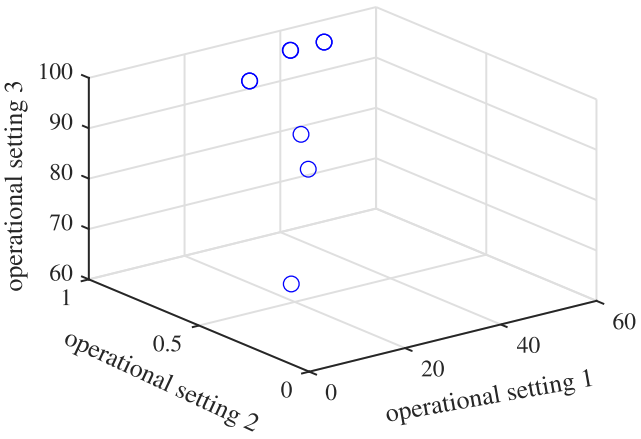


FIGURE 3. Distribution of operation modes data for all engines in the dataset.

The clustering technique can be used to find the inherent distribution structure of data and as a precursor process for learning tasks such as classification and monitoring. In this paper, the operation modes of aircraft engines are automatically divided according to the clustering result. Then, a classification model is trained based on the clustering result in the training phase to identify the operation modes of testing samples in the testing phase.

After distinguishing the operation modes, the engine data under multiple operation modes are preprocessed respectively. The process of data preprocessing mainly includes three steps, i.e., feature scaling, feature selection, and denoising processing. Feature scaling is to standardize the sensor data under each operation mode. Feature selection is to select sensors according to the trend of the collected data. The slope and changing range of sensor data are taken into consideration for selection. The sensors with consistent trends for all engines and with the close measurement values approximately when the sensor fails are selected. The top five sensors under each operation mode are selected in this paper by ranking the tendency of 21 sensors, and the sensors selected under each operation mode are different. The moving average filtering method is utilized in the step of denoising processing to smooth the sensor data so as to reduce the influence of noise.

C. ANOMALY DETECTION MODEL

In the training phase, after distinguishing the operation modes and realizing data preprocessing, the anomaly detection models under different operation modes are built. Each engine is in a healthy state at the beginning and gradually degenerates until it fails, which generates the data in the whole lifecycle of the equipment. Only in the final stage of the operation, the abnormality occurs, causing the system failure. Therefore, the data distribution of each engine is imbalanced, and the data set is composed of a large number of healthy samples and rare anomaly samples. For monitored run-to-failure systems, the data obtained in the early stage can be regarded as healthy, while those in the late stage can be regarded as abnormal [30].

In this paper, the normal model is constructed on healthy samples consisting of the first 50% of each engine from the training set. By observing the trend of the data, it is generally known that each engine begins to degrade dramatically after undergoing 80% of its life cycles and eventually fails. Therefore, this paper chooses the first 50% of each engine data in the training set as healthy samples. In fact, the percentage chosen can be adjusted in actual cases according to the specific application. After training the normal model, the corresponding anomaly indicator can be obtained for the testing samples by comparing its deviation degree from the normal model, illustrating the evolution of the degradation process for the aircraft engine.

D. PERFORMANCE EVALUATION

In order to verify the effectiveness of the proposed Multi-OC-ELM method, this paper compares the proposed method with the Single-OC-ELM method which does not distinguish the operation modes and the Multi-PCA method which distinguishes multiple operation modes. Furthermore, the simulation results of these methods are compared by using various evaluation indices. The same training set and testing set are used for both Single-OC-ELM and Multi-OC-ELM for construction and detection. Single-OC-ELM implements the data preprocessing directly for each engine and realizes the subsequent anomaly detection.

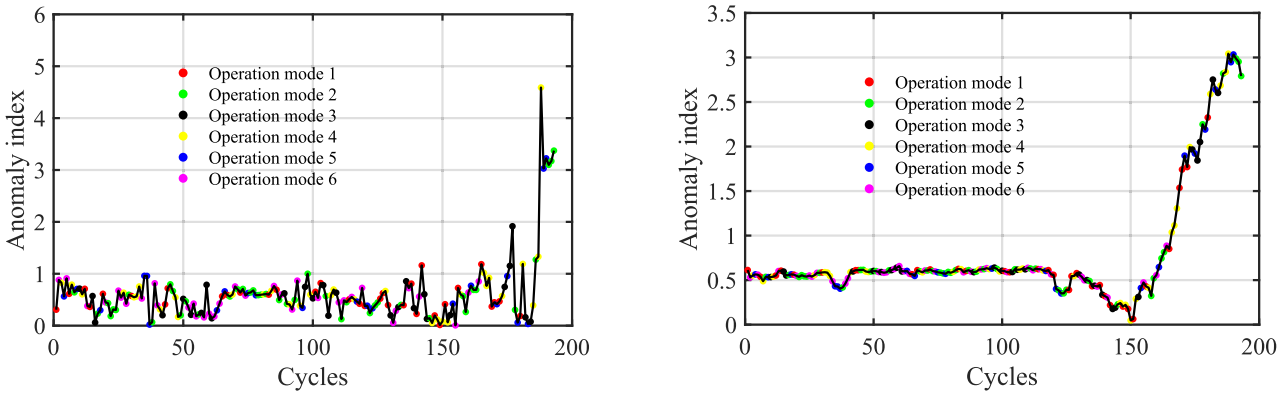
1) COMPARATIVE STUDY OF MULTI-OC-ELM AND SINGLE-OC-ELM

a: ANOMALY MONITORING INDEX

The Engine No.154 in the testing set is randomly selected to demonstrate the degradation trend of the equipment. The trends of the anomaly monitoring indices obtained by the Single-OC-ELM and the Multi-OC-ELM are presented in Fig. 4a and 4b respectively. It can be seen that in each cycle of operating, the operation modes of the engine are changed randomly. The anomaly index obtained by OC-ELM fluctuates with a smaller amplitude at the beginning and begins to decrease dramatically in the later stage, indicating that the equipment gradually deteriorates and the degree of abnormality is deepening with increasing cycles. By comparing the change of anomaly indices in Fig. 4a and 4b, apparent distinctions can be drawn that the trend of the engine degradation index obtained by Multi-OC-ELM is smoother due to the handling of the information about the operation modes, which makes the engine state of health (SoH) more distinguishable, helping to get a higher detection rate.

b: MODEL STABILITY ANALYSIS

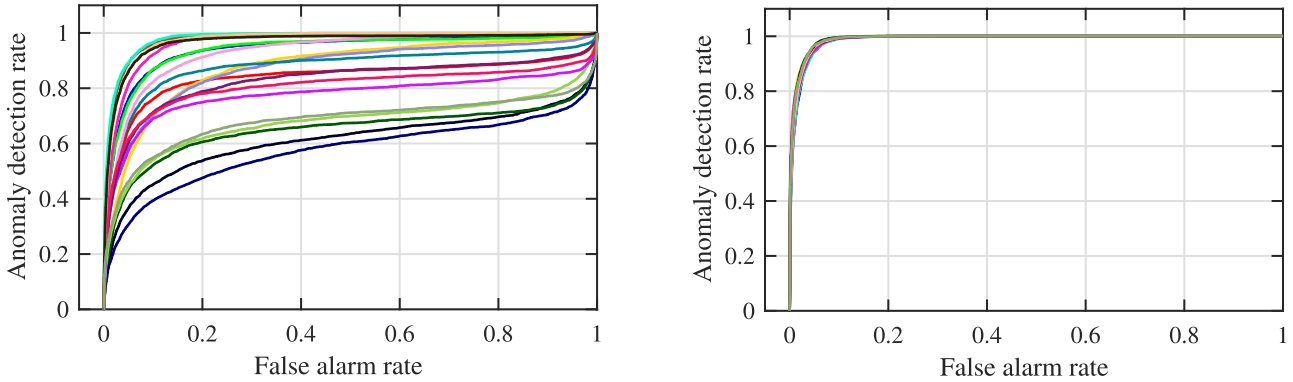
The receiver operating characteristic (ROC) curve is applied in this paper to measure the detection performance of the model [31], [32]. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR), providing tools to select possibly optimal models and to discard suboptimal ones. The anomaly indices are selected



(a) Anomaly index of Engine No.154 by using Single-OC-ELM.

(b) Anomaly index of Engine No.154 by using Multi-OC-ELM.

FIGURE 4. Anomaly index of Engine No.154 by using Single-OC-ELM and Multi-OC-ELM.



(a) Results of the Single-OC-ELM.

(b) Results of the Multi-OC-ELM.

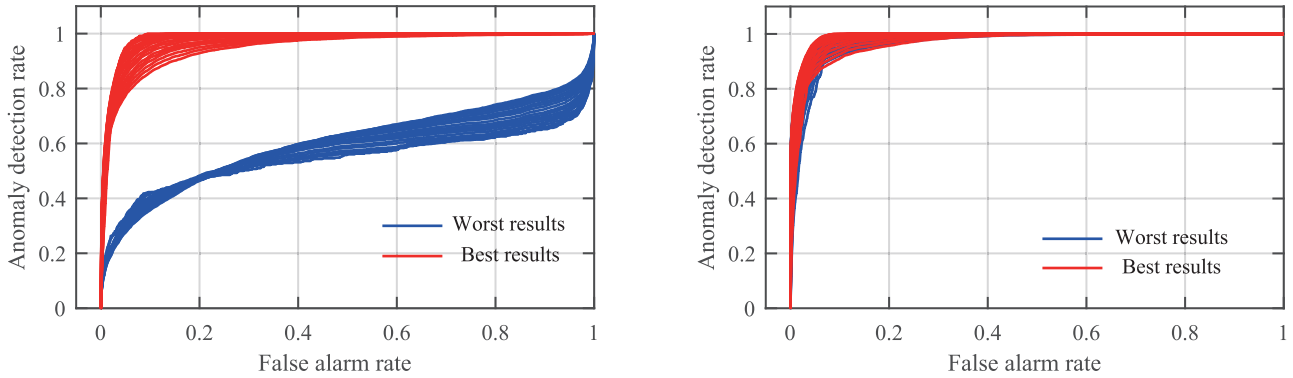
FIGURE 5. ROC curves of two methods with running 20 times repeatedly.

as the discriminant threshold of the model in proper order, forming the ROC curve. In practical application, the discriminant thresholds are determined by (13). The abscissa of the ROC curve is the false alarm rate (FAR) and the ordinate is the anomaly detection rate (ADR) which is the accuracy of the anomaly samples being detected. The closer the curve is to the upper left corner in the coordinate axis, the better the distinguishing performance of the model is. And the area under the ROC curve (AUC) provides a good reference to quantify the model. The larger AUC value indicates the stronger distinguishing capability of the model.

Since the input weights and the biases of the hidden layer in the ELM network are assigned randomly during every training process, the final output is the average value of multiple operation results. However, in the experimental simulation process, it can be noticed that there is still a great difference among the results obtained by Single-OC-ELM. The Single-OC-ELM model is unstable. Therefore, the stability of Single-OC-ELM and Multi-OC-ELM is compared in this paper.

A total of 20 ROC curves obtained by Single-OC-ELM and Multi-OC-ELM are shown in Fig. 5a and Fig. 5b, respectively, in which the simulations are independently performed 20 times and each simulation result is an average of 30 results. It can be found from the degradation trend that each engine generally begins to degrade significantly after 80% of its life cycles. Therefore, this paper labels the last 10% of each engine data as the abnormal samples. In this paper, the number of hidden layer neurons is determined by a stepwise testing method [33], [34], where the step of the method is to set an initial value first, and then the number of the hidden layer neurons increases gradually based on the initial one. The classification performance of the model is considered as the selection criteria for selecting the number of neurons. Finally, the number of the hidden layer neurons is set to 50 for Single-OC-ELM and 10 for Multi-OC-ELM. The parameter setting in this paper is a condition for the implementation of the engine case and the method illustration. In practice, the parameters can be adjusted according to application requirements.

As shown in Fig. 5a, in some cases, the AUC values of Single-OC-ELM are close to 0.5, which is equivalent to the probability of random recognition. Small AUC values mean that the abnormal samples are indistinguishable from the



(a) Results of the Single-OC-ELM method.

(b) Results of the Multi-OC-ELM method.

FIGURE 6. ROC curves of two methods under different proportions of abnormal data from 1% to 20%.

TABLE 1. Performance comparison between the Single-OC-ELM and the Multi-OC-ELM.

Method	Variance of 20 Anomaly Indexes	Variance of 20 AUC Values	Running Time (s)
Single-OC-ELM	0.397889	0.016481	3054.8887
Multi-OC-ELM	0.018846	1.767716e-06	702.8143

normal ones, resulting in poor classification effect. It can be observed from Fig. 5a and Fig. 5b that multiple ROC curves of Multi-OC-ELM are more concentrated. Furthermore, AUC of Multi-OC-ELM is larger than that of Single-OC-ELM with running 20 times. Therefore, with fewer hidden layer neurons, Multi-OC-ELM has a simpler structure and achieves better performance for anomaly recognition.

As the calculation results are shown in Table 1, the variances of the anomaly indices and the AUC obtained by Multi-OC-ELM are both much smaller than those of Single-OC-ELM with running 20 times, and the running time of Multi-OC-ELM is shorter. In conclusion, parallel monitoring based on multimodel by dividing the operation modes can not only reduce time cost, but also achieve stronger distinguishing performance and greater stability.

c: MODEL RECOGNITION CAPABILITY

To explore the influence of the labeling, the last percentages of the engine data changed from 1% to 20% are selected as anomaly samples respectively in this paper, and the corresponding ROC curves under different percentages are observed in Fig. 6. Each simulation result is the mean of 30 results. The number of neurons at the hidden layer is set to 10 for Multi-OC-ELM and 50 for Single-OC-ELM. Since the Single-OC-ELM model is unstable, the best results and the worst results by Single-OC-ELM are shown in Fig. 6a. Similarly, the corresponding results by Multi-OC-ELM are shown in Fig. 6b.

By comparing Fig. 6a and Fig. 6b, there are great differences between the best results and the worst results obtained by Single-OC-ELM, and the results of Multi-OC-ELM are more stable. Furthermore, a total of 20 ROC curves under different proportions of the abnormal data by Multi-OC-ELM are more compact than those of Single-OC-ELM. To quantify the results of Fig. 6, the corresponding AUC values are calculated in Table 2. It can be seen that the AUC values of the worst results by Multi-OC-ELM are all above 0.97, and they have a smaller fluctuating range under different proportions of the abnormal data than those of the best results by Single-OC-ELM. The minimum of the best results by Single-OC-ELM is only 0.949. Consequently, the distinguishing capability of Multi-OC-ELM is stronger, achieving good performance under different percentages of the abnormal data.

d: TYPE I AND TYPE II ERROR RATES

The last 10% data of each engine are set as the abnormal data, and then the trend of Type I and Type II error rates by the two methods can be observed in Fig. 7. Type I and Type II error rates represent the FAR and the missing alarm rate (MAR) respectively. In Fig. 7, FAR and MAR obtained by the worst results of Multi-OC-ELM versus the increasing threshold value are shown by the red lines, and the best simulation results of Single-OC-ELM are presented by the blue lines. The dot-dash lines in Fig. 7 represent MAR, while the solid ones represent FAR. As shown in Fig. 7, MAR decreases, while FAR increases with the threshold increasing. Low FAR and MAR indicate the good performance of models. However, MAR and FAR are contradictory in some cases. Therefore, the lower intersection point of the two lines indicates the better performance of the model, which means that the lower FAR and MAR can be acquired simultaneously by selecting appropriate thresholds.

In Fig. 7, it can be seen that the intersection point of red curves is significantly lower than that of the blue curves, indicating that lower FAR and MAR can be acquired by Multi-OC-ELM at the same time. In addition, compared with the red curves in Fig. 7, the two blue curves change more steeply, indicating that Single-OC-ELM has low error rates only in a small range. By contrast, the red curves

TABLE 2. Comparison between Multi-OC-ELM and Single-OC-ELM in terms of AUC.

Method	Percentage of data labeled as abnormal samples	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
Single-OC-ELM	Worst results	0.545	0.550	0.547	0.543	0.548	0.556	0.563	0.571	0.571	0.573
	Best results	0.982	0.983	0.983	0.983	0.984	0.984	0.984	0.983	0.982	0.982
Multi-OC-ELM	Worst results	0.974	0.977	0.979	0.981	0.983	0.984	0.985	0.986	0.987	0.988
	Best results	0.982	0.985	0.986	0.988	0.989	0.990	0.990	0.991	0.991	0.991
Method	Percentage of data labeled as abnormal samples	11%	12%	13%	14%	15%	16%	17%	18%	19%	20%
Single-OC-ELM	Worst results	0.577	0.582	0.589	0.591	0.591	0.595	0.600	0.600	0.601	0.601
	Best results	0.979	0.977	0.975	0.972	0.969	0.966	0.963	0.958	0.954	0.949
Multi-OC-ELM	Worst results	0.989	0.989	0.988	0.988	0.987	0.985	0.983	0.981	0.978	0.974
	Best results	0.991	0.991	0.990	0.989	0.987	0.985	0.983	0.980	0.976	0.971

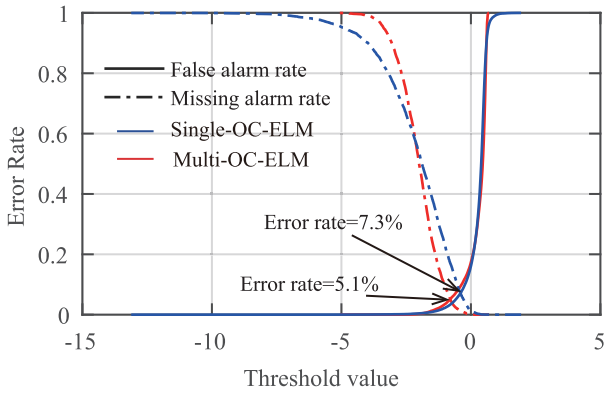


FIGURE 7. Changes of Type I and Type II error rates by Multi-OC-ELM and Single-OC-ELM.

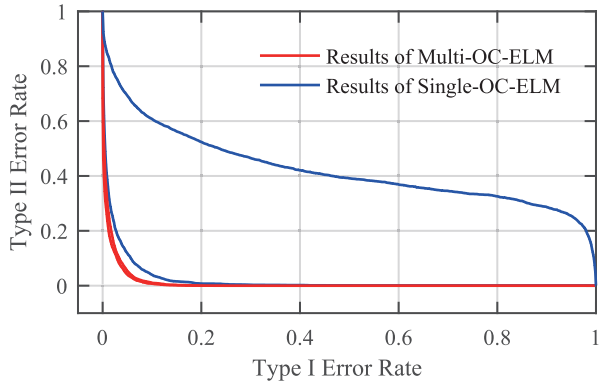


FIGURE 8. The relationship between Type I and Type II error rates obtained by Multi-OC-ELM and Single-OC-ELM.

obtained by Multi-OC-ELM vary gently, possessing a wider interval of low error rates. Therefore, low FAR and MAR can be acquired within a larger range of thresholds using Multi-OC-ELM, which can provide more choices for the user to realize alarming in the accepted time range.

In Fig. 8, the best results and the worst results of the two methods about the relationship between Type I and Type II error rates are presented. It is general that in the case of ensuring the same FAR, lower MAR is desirable, i.e., the closer the curve in Fig. 8 is to the lower-left corner, the better the model performance is. The two red curves in Fig. 8 are very close. Therefore, it can be concluded that the performance of Multi-OC-ELM is significantly better than that of Single-OC-ELM.

2) COMPARATIVE STUDY OF MULTI-OC-ELM AND MULTI-PCA

PCA is one of the most widely used multivariate statistical methods in the process monitoring of industrial systems [35]. It can detect anomalies of the equipment without the prior information of data labels, which is applicable to the industrial application. Therefore, Multi-OC-ELM is compared with Multi-PCA to illustrate the effectiveness. Note that Multi-PCA differs from Multi-OC-ELM only in the detection model.

a: ANOMALY MONITORING INDEX

The trend of the anomaly index about Engine No. 154 obtained by Multi-PCA is presented in Fig. 9. Comparing 4b with Fig. 9, it can be seen that the anomaly index of Multi-PCA fluctuates with a larger amplitude than that of Multi-OC-ELM, indicating that the index provided by the latter possesses a stronger distinguishing capability. Since the anomaly index of Multi-OC-ELM in the latter stage sharply decreases, the abnormality would be captured more accurately with a low FAR than that of Multi-PCA.

In addition, the PCA-based unsupervised algorithm for anomaly detection may lead to bad anomaly scores for the data which are actually in a healthy stage. This is caused by the instability of equipment and the lack of data at the beginning of the operation. As shown in Fig. 9, the anomaly index of Multi-PCA shows a decreasing trend and then it rises, which may result in higher false alarm rates at the early stage.

b: MODEL RECOGNITION CAPABILITY

Similarly, the different last percentage of each engine data varied from 1% to 20% are set as anomaly samples

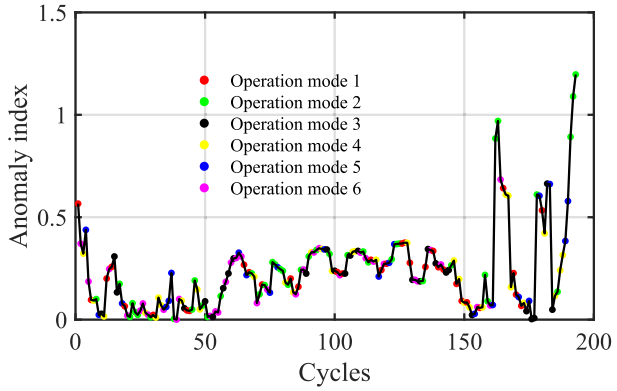


FIGURE 9. Anomaly index of Engine No.154 by using Multi-PCA.

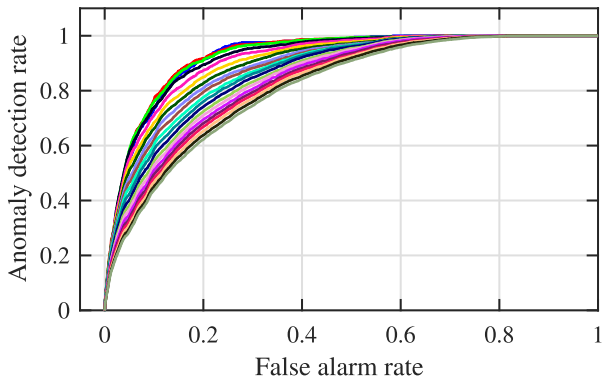


FIGURE 10. ROC curves of the Multi-PCA method under different proportions of abnormal data from 1% to 20%.

respectively, and the corresponding 20 ROC curves obtained by Multi-PCA are presented in Fig. 10. In this paper, the worst results in 20 simulations by Multi-OC-ELM are compared with the result of Multi-PCA. Comparing Fig. 6b with Fig. 10, it can be seen that the 20 ROC curves obtained by Multi-OC-ELM are more concentrated, while the variation is larger for the case of Multi-PCA. In Fig. 11, the AUC values obtained by Multi-OC-ELM under different percentages are all above 0.97, while the largest AUC value obtained by Multi-PCA is around 0.92. In addition, all the AUC values obtained by Multi-PCA in each case are lower than those of Multi-OC-ELM. As the percentage increases, the AUC value obtained by Multi-PCA decreases significantly, indicating that the distinguishing capability gradually weakens when the threshold is far away from failure point. It can be concluded that the performance of Multi-OC-ELM is less affected by the labeling of the anomaly data, representing a stronger robustness of Multi-OC-ELM.

The confusion matrices obtained by Multi-OC-ELM and Multi-PCA are presented in Fig. 12. Each row of the matrix represents the instances in a predicted health state, while each column represents the instances in an actual health state. In Fig. 12, the value 1 represents the healthy class and value 0 represents the anomaly class. The green areas and red areas represent the number of samples and the percentage of the total samples. The green areas correspond to the correctly

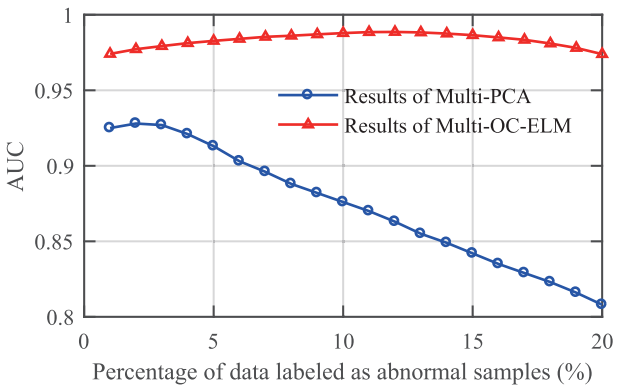


FIGURE 11. Comparison of AUC values by the two methods under different percentages.

classified samples and the red areas correspond to the incorrectly classified samples. A column of gray areas on the right of the confusion matrix correspond to the precision for each class, in which the precision is the accuracy rate in the predicted samples. The gray areas at the bottom row correspond to the true positive rate (TPR) which is the percentage of the correctly classified samples in each class. The blue area in the lower right corner represents the overall accuracy. It can be observed clearly in Fig. 12 that each metric in the confusion matrix obtained by Multi-OC-ELM performs better than that of Multi-PCA.

c: MODEL STABILITY ANALYSIS

PCA constructs the detection model merely based on the data, which indicates that the model is constant. The ROC curve obtained by Multi-PCA and a total of 20 ROC curves obtained by Multi-OC-ELM with simulating 20 times are shown in Fig. 13. The last 10% of each engine data are set as the anomaly samples, which is a condition chosen to accomplish the comparison between Multi-OC-ELM and Multi-PCA on the stability analysis.

As shown in Fig. 13, the AUC value of Multi-PCA is lower than any result of Multi-OC-ELM. Generally, since the OC-ELM algorithm involves label information, the Multi-OC-ELM method has a better performance of anomaly detection.

d: TYPE I AND TYPE II ERROR RATES

As the threshold increases, the changes of FAR and MAR obtained by Multi-PCA are plotted in Fig. 14. Comparing Fig. 14 with Fig. 7, the intersection point of FAR and MAR obtained by Multi-OC-ELM is lower than that of Multi-PCA, indicating that it is possible for Multi-OC-ELM to simultaneously obtain a FAR and a MAR less than 5% by selecting an appropriate threshold, which is impossible for Multi-PCA. Moreover, the evolution of two curves of Multi-OC-ELM in Fig. 7 is smoother, indicating that low MAR and FAR can be acquired within a wider range of thresholds. Fig. 15 has the same form as Fig. 8, and it can be concluded that the worst

(a) Confusion matrix obtained by the Multi-OC-ELM method.

(b) Confusion matrix obtained by the Multi-PCA method.

FIGURE 12. Confusion matrix obtained by the two methods.

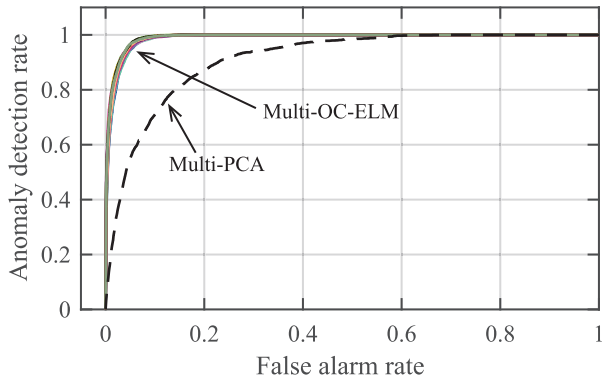


FIGURE 13. ROC curves of the Multi-PCA method and the Multi-OC-ELM method with running 20 times repeatedly.

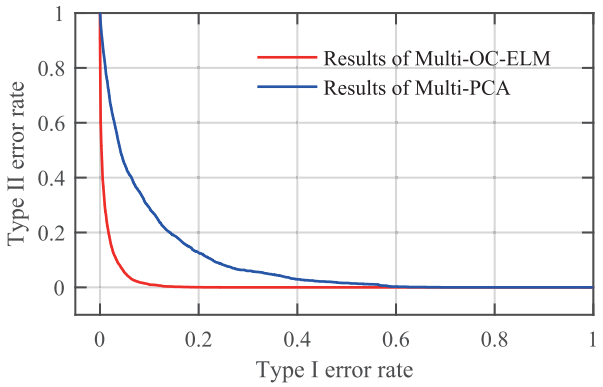


FIGURE 15. The relationship between Type I and Type II error rates obtained by Multi-OC-ELM and Multi-PCA.

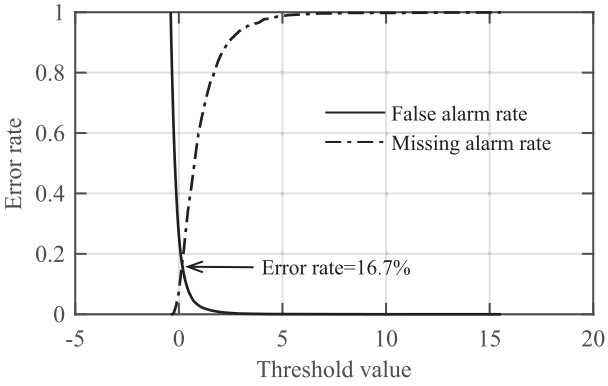


FIGURE 14. Changes of Type I and Type II error rates by Multi-PCA.

result of Multi-OC-ELM performs significantly better than that of Multi-PCA.

In this paper, the proposed method is compared with Single-OC-ELM and Multi-PCA respectively. It is significant to establish multiple models by distinguishing multiple operation modes and the effectiveness of OC-ELM is verified. In addition, different general evaluation indicators such as ROC, Type I and Type II error rates, confusion matrix, and running time are applied for a comprehensive comparison. It is demonstrated that the proposed method performs better with larger AUC values and lower error rates. Since the proposed method is based on ELM, its performance can be closely related to the number of the hidden layer neurons, or the complexity of the structures of the distributed data. For the problems where the data are distributed in a complex nonlinear way, increasing the number of the hidden layer neurons and combining it with effective feature extraction methods can be possible solutions.

IV. CONCLUSION

In this paper, a new anomaly detection framework for complex industrial systems with multiple operation modes is proposed. By distinguishing different operation modes and constructing multiple models, the proposed method can capture the anomaly changes of the system accurately under multiple operation modes and verified by a field dataset C-MAPSS of the aircraft engine. OC-ELM is utilized for anomaly detection, which is applicable for the practical industry process with rare anomaly samples. The effectiveness and the superiority of the proposed method are

demonstrated with multiple performance metrics compared with the existing methods. The anomaly index obtained by the detection model effectively reflects the change of the degradation trend of the engine, so as to detect abnormalities of the system in time. The proposed method has superior performance in robustness and practicability without the parameter optimization, which is beneficial to the multimode anomaly detection in most industrial systems.

REFERENCES

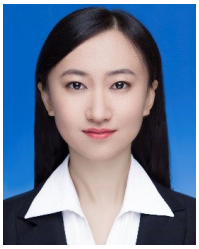
[1] S. Zhao, V. Makis, S. Chen, and Y. Li, "Health assessment method for electronic components subject to condition monitoring and hard failure," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 1, pp. 138–150, Jan. 2019.
[2] S. Zhao, V. Makis, S. Chen, and Y. Li, "Evaluation of reliability function and mean residual life for degrading systems subject to condition monitoring and random failure," *IEEE Trans. Rel.*, vol. 67, no. 1, pp. 13–25, Mar. 2018.
[3] W. Jiang, Y. Hong, B. Zhou, X. He, and C. Cheng, "A GAN-based anomaly detection approach for imbalanced industrial time series," *IEEE Access*, vol. 7, pp. 143608–143619, 2019.
[4] C. Yang, J. Liu, Y. Zeng, and G. Xie, "Real-time condition monitoring and fault detection of components based on machine-learning reconstruction model," *Renew. Energy*, vol. 133, pp. 433–441, Apr. 2019.
[5] H. Zhang, X. Deng, Y. Zhang, C. Hou, C. Li, and Z. Xin, "Nonlinear process monitoring based on global preserving unsupervised kernel extreme learning machine," *IEEE Access*, vol. 7, pp. 106053–106064, 2019.
[6] C. Yang, L. Zhou, K. Huang, H. Ji, C. Long, X. Chen, and Y. Xie, "Multimode process monitoring based on robust dictionary learning with application to aluminium electrolysis process," *Neurocomputing*, vol. 332, pp. 305–319, Mar. 2019.
[7] Y. Xu, S.-Q. Shen, Y.-L. He, and Q.-X. Zhu, "A novel hybrid method integrating ICA-PCA with relevant vector machine for multivariate process monitoring," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 4, pp. 1780–1787, Jul. 2019.
[8] F. Wang, S. Zhang, and Y. Yin, "Log likelihood monitoring for multimode process using variational Bayesian mixture factor analysis model," *IEEE Access*, vol. 7, pp. 89083–89092, 2019.
[9] Y. Wang, F. Sun, and B. Li, "Multiscale neighborhood normalization-based multiple dynamic PCA monitoring method for batch processes with frequent operations," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 3, pp. 1053–1064, Jul. 2018.
[10] Q. Jiang, X. Yan, and B. Huang, "Performance-driven distributed PCA process monitoring based on fault-relevant variable selection and Bayesian inference," *IEEE Trans. Ind. Electron.*, vol. 63, no. 1, pp. 377–386, Jan. 2016.
[11] S. Zhao, S. Chen, F. Yang, E. Ugur, B. Akin, and H. Wang, "A composite failure precursor for condition monitoring and remaining useful life prediction of discrete power devices," *IEEE Trans. Ind. Informat.*, vol. 17, no. 1, pp. 688–698, Jan. 2021.
[12] T. Sukchotrat, S. B. Kim, and F. Tsung, "One-class classification-based control charts for multivariate process monitoring," *IIE Trans.*, vol. 42, no. 2, pp. 107–120, Nov. 2009.
[13] Q. Leng, H. Qi, J. Miao, W. Zhu, and G. Su, "One-class classification with extreme learning machine," *Math. Problems Eng.*, vol. 2015, pp. 1–11, May 2015.
[14] V. M. Janakiraman and D. Nielsen, "Anomaly detection in aviation data using extreme learning machines," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 1993–2000.
[15] Y. Imamverdiyev and L. Sukhostat, "Anomaly detection in network traffic using extreme learning machine," in *Proc. IEEE 10th Int. Conf. Appl. Inf. Commun. Technol. (AICT)*, Oct. 2016, pp. 1–4.
[16] W. Yan, "One-class extreme learning machines for gas turbine combustor anomaly detection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 2909–2914.
[17] V. Mygdalis, A. Iosifidis, A. Tefas, and I. Pitas, "Laplacian one class extreme learning machines for human action recognition," in *Proc. IEEE 18th Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2016, pp. 1–5.
[18] S. Wang, E. Zhu, and J. Yin, "Video anomaly detection based on ULGP-OF descriptor and one-class ELM," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 2630–2637.
[19] W. Yan, L. K. Mestha, and M. Abbaszadeh, "Attack detection for securing cyber physical systems," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8471–8481, Oct. 2019.
[20] P. Wang, Z. Long, and N. Dang, "Multi-model switching based fault detection for the suspension system of maglev train," *IEEE Access*, vol. 7, pp. 6831–6841, 2019.
[21] A. Bakdi, A. Kouadri, and S. Mekhilef, "A data-driven algorithm for online detection of component and system faults in modern wind turbines at different operating zones," *Renew. Sustain. Energy Rev.*, vol. 103, pp. 546–555, Apr. 2019.
[22] L. Wang, X. Deng, and Y. Cao, "Multimode complex process monitoring using double-level local information based local outlier factor method," *J. Chemometrics*, vol. 32, no. 10, p. e3048, Oct. 2018.
[23] B. Wang, Z. Li, and X. Yan, "Multi-subspace factor analysis integrated with support vector data description for multimode process monitoring," *J. Franklin Inst.*, vol. 355, no. 15, pp. 7664–7690, Oct. 2018.
[24] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.
[25] P. Wen, S. Zhao, S. Chen, and Y. Li, "A generalized remaining useful life prediction method for complex systems based on composite health indicator," *Rel. Eng. Syst. Saf.*, vol. 205, Jan. 2021, Art. no. 107241.
[26] *MATLAB 2016B*. Mathworks. Accessed: Dec. 2020. [Online]. Available: https://www.mathworks.com/
[27] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *Proc. Int. Conf. Prognostics Health Manage.*, Denver, CO, USA, Oct. 2008, pp. 1–9.
[28] *C-MAPSS Data Set*. NASA PCoE. Accessed: Jul. 2018. [Online]. Available: https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository
[29] W. Yu, I. Y. Kim, and C. Mechefske, "Remaining useful life estimation using a bidirectional recurrent neural network based autoencoder scheme," *Mech. Syst. Signal Process.*, vol. 129, pp. 764–780, Aug. 2019.
[30] D. Wang, Z. Peng, and L. Xi, "The sum of weighted normalized square envelope: A unified framework for kurtosis, negative entropy, Gini index and smoothness index for machine health monitoring," *Mech. Syst. Signal Process.*, vol. 140, Jun. 2020, Art. no. 106725.
[31] B. Siegel, "Industrial anomaly detection: A comparison of unsupervised neural network architectures," *IEEE Sensors Lett.*, vol. 4, no. 8, pp. 1–4, Aug. 2020.
[32] S. Zavrak and M. Iskefiyeli, "Anomaly-based intrusion detection from network flow features using variational autoencoder," *IEEE Access*, vol. 8, pp. 108346–108358, 2020.
[33] F. Huang, J. Lu, J. Tao, L. Li, X. Tan, and P. Liu, "Research on optimization methods of ELM classification algorithm for hyperspectral remote sensing images," *IEEE Access*, vol. 7, pp. 108070–108089, 2019.
[34] D. Xue, X. Jing, and H. Liu, "Detection of false data injection attacks in smart grid utilizing ELM-based OCON framework," *IEEE Access*, vol. 7, pp. 31762–31773, 2019.
[35] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6418–6428, Nov. 2014.

SHAOWEI CHEN (Member, IEEE) is currently an Associate Professor with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China, where he is also the Dean of the Department of Communication Engineering and the Director of Signal Processing and Control Technology Laboratory. He is the Principal Investigator of several projects supported by the Aeronautical Science Foundation of China, Beijing, China. His research interests include fault diagnosis, sensors, condition monitoring, and prognosis of electronic systems. He is a Senior Member of the Chinese Institute of Electronics. He is selected to receive several provincial and ministerial science and technology awards and has five patents.

MENG WU received the B.S. degree in communication engineering from Ningxia University, Yinchuan, China, in 2017. She is currently pursuing the M.S. degree in communication and information system with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China. Her research interests include the anomaly detection of complex industrial processes, data analytics, and machine learning algorithm.

PENGFEI WEN (Graduate Student Member, IEEE) received the B.S. degree in communication engineering from the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China, in 2017, where he is currently pursuing the Ph.D. degree in information and communication engineering. His research interests include statistical modeling, data and information fusion for data analytics, and prognostics.

FANGDA XU received the B.S. degree in communication engineering from the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China, in 2019, where she is currently pursuing the M.S. degree in communication and information system. Her research interest includes the time series anomaly detection.



SHENGYUE WANG received the B.S. degree in electronic science and technology from Northeastern University, Shenyang, China, in 2019. She is currently pursuing the M.S. degree in communication and information system with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China. Her research interests include the reliability of power electronic converter systems and machine learning.

SHUAI ZHAO (Member, IEEE) received the B.S. degree in communication engineering, the M.S. degree in communication and information system, and the Ph.D. degree in information and communication engineering from Northwestern Polytechnical University, Xi'an, China, in 2011, 2014, and 2018, respectively.

From 2014 to 2016, he was a Visiting Ph.D. Student with the Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada, with the scholarship from the China Scholarship Council (CSC). In August 2018, he was a Visiting Scholar with the Power Electronics and Drives Laboratory, Department of Electrical and Computer Science, The University of Texas at Dallas, Richardson, TX, USA. He is currently a Postdoctoral Researcher with the Center of Reliable Power Electronics (CORPE), Department of Energy Technology, Aalborg University, Aalborg, Denmark. His research interests include condition monitoring, data analytics, residual life prediction, and health assessment of power electronic systems.